

PCT09

RAW SEQUENCE LISTING

DATE: 05/23/2001

PATENT APPLICATION: US/09/831,180

TIME: 16:02:14

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

4 <110> APPLICANT: Chiaki Senoo et al.
6 <120> TITLE OF INVENTION: Novel Trypsin Family Serine Proteases
9 <130> FILE REFERENCE: 50026/027001
C--> 11 <140> CURRENT APPLICATION NUMBER: US/09/831,180
C--> 11 <141> CURRENT FILING DATE: 2001-05-03
11 <150> PRIOR APPLICATION NUMBER: JP 1998-313366
12 <151> PRIOR FILING DATE: 1998-11-04
14 <160> NUMBER OF SEQ ID NOS: 45
16 <170> SOFTWARE: FastSEQ for Windows Version 4.0
18 <210> SEQ ID NO: 1
19 <211> LENGTH: 1033
20 <212> TYPE: DNA
21 <213> ORGANISM: Mus musculus
23 <220> FEATURE:
24 <221> NAME/KEY: CDS
25 <222> LOCATION: (48)...(1010)
27 <400> SEQUENCE: 1
28 cctgcctcag tggttgagct cccattgct gatgtgcagg caagccg atg aaa cga 56
29 Met Lys Arg
30 1
32 tgg aag gac aga aga aca ggc ctg ttg ctg cca ttg gtc ctc ctg ttg 104
33 Trp Lys Asp Arg Arg Thr Gly Leu Leu Leu Pro Leu Val Leu Leu Leu
34 5 10 15
36 ttt ggg gca tgt agc tca ctg gca tgg gta tgt ggc cgg cga atg agt 152
37 Phe Gly Ala Cys Ser Ser Leu Ala Trp Val Cys Gly Arg Arg Met Ser
38 20 25 30 35
40 agc aga tcc caa caa ctt aac aat gct tct gct atc gtg gaa ggc aaa 200
41 Ser Arg Ser Gln Gln Leu Asn Asn Ala Ser Ala Ile Val Glu Gly Lys
42 40 45 50
44 cct gct tct gct atc gtg gga ggc aaa cct gca aac atc ttg gag ttc 248
45 Pro Ala Ser Ala Ile Val Gly Gly Lys Pro Ala Asn Ile Leu Glu Phe
46 55 60 65
48 ccc tgg cat gtg ggg att atg aat cat ggt agt cat ctc tgt ggg gga 296
49 Pro Trp His Val Gly Ile Met Asn His Gly Ser His Leu Cys Gly Gly
50 70 75 80
52 tct att ctc aat gag tgg tgg gtt cta tct gca tcc cat tgc ttc gac 344
53 Ser Ile Leu Asn Glu Trp Trp Val Leu Ser Ala Ser His Cys Phe Asp
54 85 90 95
56 caa cta aac aac tct aaa ttg gag atc att cat ggc act gaa gac ctc 392
57 Gln Leu Asn Asn Ser Lys Leu Glu Ile Ile His Gly Thr Glu Asp Leu
58 100 105 110 115
60 agc aca aag ggc ata aag tat cag aaa gtg gac aag tta ttc ttg cac 440
61 Ser Thr Lys Gly Ile Lys Tyr Gln Lys Val Asp Lys Leu Phe Leu His
62 120 125 130
64 cca aag ttt gat gac tgg ctc ctg gac aac gac ata gct ttg ctc ttg 488
65 Pro Lys Phe Asp Asp Trp Leu Leu Asp Asn Asp Ile Ala Leu Leu Leu
66 135 140 145

ENTERED

RAW SEQUENCE LISTING

DATE: 05/23/2001

PATENT APPLICATION: US/09/831,180

TIME: 16:02:14

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

```

68 ctc aaa tcc cca tta aac ttg agt gtc aac agg ata cct atc tgc act 536
69 Leu Lys Ser Pro Leu Asn Leu Ser Val Asn Arg Ile Pro Ile Cys Thr
70      150      155      160
72 tca gaa atc tct gac ata cag gca tgg agg aac tgc tgg gtg aca gga 584
73 Ser Glu Ile Ser Asp Ile Gln Ala Trp Arg Asn Cys Trp Val Thr Gly
74      165      170      175
76 tgg ggc att act aat act agt gaa aaa gga gtc caa ccc aca att ctg 632
77 Trp Gly Ile Thr Asn Thr Ser Glu Lys Gly Val Gln Pro Thr Ile Leu
78 180      185      190      195
80 cag gca gtc aaa gtg gat ctg tac aga tgg gat tgg tgt ggc tat att 680
81 Gln Ala Val Lys Val Asp Leu Tyr Arg Trp Asp Trp Cys Gly Tyr Ile
82      200      205      210
84 ttg tct cta tta acc aag aat atg ctg tgt gct ggg act caa gat cct 728
85 Leu Ser Leu Leu Thr Lys Asn Met Leu Cys Ala Gly Thr Gln Asp Pro
86      215      220      225
88 ggg aag gat gcc tgc cag ggc gac agt gga gga gct ctc gtt tgc aac 776
89 Gly Lys Asp Ala Cys Gln Gly Asp Ser Gly Gly Ala Leu Val Cys Asn
90      230      235      240
92 aaa aag aga aac aca gcc att tgg tac cag gtg ggc att gtc agc tgg 824
93 Lys Lys Arg Asn Thr Ala Ile Trp Tyr Gln Val Gly Ile Val Ser Trp
94      245      250      255
96 ggc atg ggc tgt ggc aag aag aat ctg cca gga gta tac acc aag gtg 872
97 Gly Met Gly Cys Gly Lys Lys Asn Leu Pro Gly Val Tyr Thr Lys Val
98 260      265      270      275
100 tca cac tat gtg agg tgg atc agc aag cag aca gcg aag gcg ggg agg 920
101 Ser His Tyr Val Arg Trp Ile Ser Lys Gln Thr Ala Lys Ala Gly Arg
102      280      285      290
104 cct tat atg tat gag cag aac tct gcg tgc cct ttg gtg ctc tct tgc 968
105 Pro Tyr Met Tyr Glu Gln Asn Ser Ala Cys Pro Leu Val Leu Ser Cys
106      295      300      305
108 cgg gct atc ttg ttc cta tat ttt gta atg ttt ctt cta acc 1010
109 Arg Ala Ile Leu Phe Leu Tyr Phe Val Met Phe Leu Leu Thr
110      310      315      320
112 tgatgattaa acgtgagact gcc 1033
114 <210> SEQ ID NO: 2
115 <211> LENGTH: 321
116 <212> TYPE: PRT
117 <213> ORGANISM: Mus musculus
119 <400> SEQUENCE: 2
120 Met Lys Arg Trp Lys Asp Arg Arg Thr Gly Leu Leu Leu Pro Leu Val
121 1      5      10      15
122 Leu Leu Leu Phe Gly Ala Cys Ser Ser Leu Ala Trp Val Cys Gly Arg
123      20      25      30
124 Arg Met Ser Ser Arg Ser Gln Gln Leu Asn Asn Ala Ser Ala Ile Val
125      35      40      45
126 Glu Gly Lys Pro Ala Ser Ala Ile Val Gly Gly Lys Pro Ala Asn Ile
127      50      55      60
128 Leu Glu Phe Pro Trp His Val Gly Ile Met Asn His Gly Ser His Leu
129 65      70      75      80

```

RAW SEQUENCE LISTING

DATE: 05/23/2001

PATENT APPLICATION: US/09/831,180

TIME: 16:02:14

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

```

130 Cys Gly Gly Ser Ile Leu Asn Glu Trp Trp Val Leu Ser Ala Ser His
131      85      90      95
132 Cys Phe Asp Gln Leu Asn Asn Ser Lys Leu Glu Ile Ile His Gly Thr
133      100      105      110
134 Glu Asp Leu Ser Thr Lys Gly Ile Lys Tyr Gln Lys Val Asp Lys Leu
135      115      120      125
136 Phe Leu His Pro Lys Phe Asp Asp Trp Leu Leu Asp Asn Asp Ile Ala
137      130      135      140
138 Leu Leu Leu Leu Lys Ser Pro Leu Asn Leu Ser Val Asn Arg Ile Pro
139 145      150      155      160
140 Ile Cys Thr Ser Glu Ile Ser Asp Ile Gln Ala Trp Arg Asn Cys Trp
141      165      170      175
142 Val Thr Gly Trp Gly Ile Thr Asn Thr Ser Glu Lys Gly Val Gln Pro
143      180      185      190
144 Thr Ile Leu Gln Ala Val Lys Val Asp Leu Tyr Arg Trp Asp Trp Cys
145      195      200      205
146 Gly Tyr Ile Leu Ser Leu Leu Thr Lys Asn Met Leu Cys Ala Gly Thr
147      210      215      220
148 Gln Asp Pro Gly Lys Asp Ala Cys Gln Gly Asp Ser Gly Gly Ala Leu
149 225      230      235      240
150 Val Cys Asn Lys Lys Arg Asn Thr Ala Ile Trp Tyr Gln Val Gly Ile
151      245      250      255
152 Val Ser Trp Gly Met Gly Cys Gly Lys Asn Leu Pro Gly Val Tyr
153      260      265      270
154 Thr Lys Val Ser His Tyr Val Arg Trp Ile Ser Lys Gln Thr Ala Lys
155      275      280      285
156 Ala Gly Arg Pro Tyr Met Tyr Glu Gln Asn Ser Ala Cys Pro Leu Val
157      290      295      300
158 Leu Ser Cys Arg Ala Ile Leu Phe Leu Tyr Phe Val Met Phe Leu Leu
159 305      310      315      320
160 Thr

```

164 <210> SEQ ID NO: 3

165 <211> LENGTH: 1034

166 <212> TYPE: DNA

167 <213> ORGANISM: Mus musculus

169 <220> FEATURE:

170 <221> NAME/KEY: CDS

171 <222> LOCATION: (69)...(1025)

172 <223> OTHER INFORMATION:

174 <221> NAME/KEY: misc_feature

175 <222> LOCATION: 10

176 <223> OTHER INFORMATION: n = A or C or G or T/U

178 <400> SEQUENCE: 3

```

W--> 179 cccacgcgtn'cggttgatc aatgtgggca gggcatcaag gcaggcacca ctgcactgga 60
180 atgacaac atg atg ctc cca ctt cta att gca ctg ctc atg gct tcc aag 110
181      Met Met Leu Pro Leu Leu Ile Ala Leu Leu Met Ala Ser Lys
182      1      5      10
184 gga caa gct aag gac cag caa gaa tca gtt ctg tgt ggc cac aga cct 158
185 Gly Gln Ala Lys Asp Gln Gln Glu Ser Val Leu Cys Gly His Arg Pro

```

RAW SEQUENCE LISTING

DATE: 05/23/2001

PATENT APPLICATION: US/09/831,180

TIME: 16:02:14

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

186	15				20				25				30				
188	gcc	ttc	cca	aac	tca	tca	tgg	ctg	cca	ttg	cgg	gag	ctg	ctt	gag	gtc	206
189	Ala	Phe	Pro	Asn	Ser	Ser	Trp	Leu	Pro	Leu	Arg	Glu	Leu	Leu	Glu	Val	
190				35					40				45				
192	cag	cat	ggg	gag	ttc	cca	tgg	caa	gtg	agt	atc	cag	atg	ctt	ggg	aaa	254
193	Gln	His	Gly	Glu	Phe	Pro	Trp	Gln	Val	Ser	Ile	Gln	Met	Leu	Gly	Lys	
194				50				55					60				
196	cac	ctg	tgt	gga	ggc	tcc	atc	atc	cac	cgg	tgg	tgg	gtt	ctg	aca	gca	302
197	His	Leu	Cys	Gly	Gly	Ser	Ile	Ile	His	Arg	Trp	Trp	Val	Leu	Thr	Ala	
198			65				70						75				
200	gca	cac	tgc	ttc	ccg	aga	acc	cta	tta	gaa	ctg	gta	gca	gtc	aat	gtc	350
201	Ala	His	Cys	Phe	Pro	Arg	Thr	Leu	Leu	Glu	Leu	Val	Ala	Val	Asn	Val	
202		80					85					90					
204	act	gtg	gtc	atg	gga	atc	aag	act	ttc	agt	gac	acc	aac	tta	gag	aga	398
205	Thr	Val	Val	Met	Gly	Ile	Lys	Thr	Phe	Ser	Asp	Thr	Asn	Leu	Glu	Arg	
206	95				100				105				110				
208	aaa	caa	gtg	cag	aag	atc	att	gct	cac	aga	gac	tac	aaa	ccg	ccc	gac	446
209	Lys	Gln	Val	Gln	Lys	Ile	Ile	Ala	His	Arg	Asp	Tyr	Lys	Pro	Pro	Asp	
210				115				120					125				
212	ctt	gac	agc	gac	ctc	tgc	ctg	ctc	cta	ctt	gcc	acg	cca	atc	caa	ttc	494
213	Leu	Asp	Ser	Asp	Leu	Cys	Leu	Leu	Leu	Ala	Thr	Pro	Ile	Gln	Phe		
214				130				135					140				
216	aat	aaa	gac	aaa	atg	ccc	atc	tgc	ctg	cca	cag	agg	gag	aac	tcc	tgg	542
217	Asn	Lys	Asp	Lys	Met	Pro	Ile	Cys	Leu	Pro	Gln	Arg	Glu	Asn	Ser	Trp	
218		145					150					155					
220	gac	cgg	tgc	tgg	atg	tca	gag	tgg	gca	tat	act	cat	ggc	cat	ggg	tca	590
221	Asp	Arg	Cys	Trp	Met	Ser	Glu	Trp	Ala	Tyr	Thr	His	Gly	His	Gly	Ser	
222		160					165					170					
224	gcc	aaa	ggc	tca	aac	atg	cac	ctg	aag	aag	ctc	agg	gtg	gtt	cag	att	638
225	Ala	Lys	Gly	Ser	Asn	Met	His	Leu	Lys	Lys	Leu	Arg	Val	Val	Gln	Ile	
226	175				180				185				190				
228	agc	tgg	agg	aca	tgt	gcg	aag	agg	gtg	act	cag	ctc	tcc	agg	aac	atg	686
229	Ser	Trp	Arg	Thr	Cys	Ala	Lys	Arg	Val	Thr	Gln	Leu	Ser	Arg	Asn	Met	
230				195				200				205					
232	ctt	tgt	gct	tgg	aag	gaa	gtg	ggc	acc	aac	ggc	aag	tgc	cag	gga	gac	734
233	Leu	Cys	Ala	Trp	Lys	Glu	Val	Gly	Thr	Asn	Gly	Lys	Cys	Gln	Gly	Asp	
234			210					215				220					
236	agc	ggg	gca	ccc	atg	gtc	tgt	gct	aac	tgg	gag	act	cgg	aga	ctc	ttt	782
237	Ser	Gly	Ala	Pro	Met	Val	Cys	Ala	Asn	Trp	Glu	Thr	Arg	Arg	Leu	Phe	
238			225					230				235					
240	caa	gtg	ggg	gtc	ttc	agc	tgg	ggc	ata	act	tca	gga	tcc	agg	ggg	agg	830
241	Gln	Val	Gly	Val	Phe	Ser	Trp	Gly	Ile	Thr	Ser	Gly	Ser	Arg	Gly	Arg	
242		240					245					250					
244	cca	ggc	att	ttt	gtg	tct	gtg	gct	cag	ttt	atc	cca	tgg	atc	ctg	gag	878
245	Pro	Gly	Ile	Phe	Val	Ser	Val	Ala	Gln	Phe	Ile	Pro	Trp	Ile	Leu	Glu	
246	255				260			265				270					
248	gag	aca	caa	agg	gag	gga	cga	gcc	ctt	gcc	ctc	tca	aag	gcc	tca	aaa	926
249	Glu	Thr	Gln	Arg	Glu	Gly	Arg	Ala	Leu	Ala	Leu	Ser	Lys	Ala	Ser	Lys	
250				275				280				285					

RAW SEQUENCE LISTING

DATE: 05/23/2001

PATENT APPLICATION: US/09/831,180

TIME: 16:02:14

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

```

252 agt ctc ttg gct ggc agt cca cgc tac cat ccc ata ttg cta agc atg 974
253 Ser Leu Leu Ala Gly Ser Pro Arg Tyr His Pro Ile Leu Leu Ser Met
254          290          295          300
256 ggc tct caa ata ctg ctt gct gcc ata ttt tct gat gat aaa tca aat 1022
257 Gly Ser Gln Ile Leu Leu Ala Ala Ile Phe Ser Asp Asp Lys Ser Asn
258          305          310          315
260 tgc/taagctctg 1034
261 Cys
265 <210> SEQ ID NO: 4
266 <211> LENGTH: 319
267 <212> TYPE: PRT
268 <213> ORGANISM: Mus musculus
270 <400> SEQUENCE: 4
271 Met Met Leu Pro Leu Leu Ile Ala Leu Leu Met Ala Ser Lys Gly Gln
272 1 5 10 15
273 Ala Lys Asp Gln Gln Glu Ser Val Leu Cys Gly His Arg Pro Ala Phe
274 20 25 30
275 Pro Asn Ser Ser Trp Leu Pro Leu Arg Glu Leu Leu Glu Val Gln His
276 35 40 45
277 Gly Glu Phe Pro Trp Gln Val Ser Ile Gln Met Leu Gly Lys His Leu
278 50 55 60
279 Cys Gly Gly Ser Ile Ile His Arg Trp Trp Val Leu Thr Ala Ala His
280 65 70 75 80
281 Cys Phe Pro Arg Thr Leu Leu Glu Leu Val Ala Val Asn Val Thr Val
282 85 90 95
283 Val Met Gly Ile Lys Thr Phe Ser Asp Thr Asn Leu Glu Arg Lys Gln
284 100 105 110
285 Val Gln Lys Ile Ile Ala His Arg Asp Tyr Lys Pro Pro Asp Leu Asp
286 115 120 125
287 Ser Asp Leu Cys Leu Leu Leu Leu Ala Thr Pro Ile Gln Phe Asn Lys
288 130 135 140
289 Asp Lys Met Pro Ile Cys Leu Pro Gln Arg Glu Asn Ser Trp Asp Arg
290 145 150 155 160
291 Cys Trp Met Ser Glu Trp Ala Tyr Thr His Gly His Gly Ser Ala Lys
292 165 170 175
293 Gly Ser Asn Met His Leu Lys Lys Leu Arg Val Val Gln Ile Ser Trp
294 180 185 190
295 Arg Thr Cys Ala Lys Arg Val Thr Gln Leu Ser Arg Asn Met Leu Cys
296 195 200 205
297 Ala Trp Lys Glu Val Gly Thr Asn Gly Lys Cys Gln Gly Asp Ser Gly
298 210 215 220
299 Ala Pro Met Val Cys Ala Asn Trp Glu Thr Arg Arg Leu Phe Gln Val
300 225 230 235 240
301 Gly Val Phe Ser Trp Gly Ile Thr Ser Gly Ser Arg Gly Arg Pro Gly
302 245 250 255
303 Ile Phe Val Ser Val Ala Gln Phe Ile Pro Trp Ile Leu Glu Glu Thr
304 260 265 270
305 Gln Arg Glu Gly Arg Ala Leu Ala Leu Ser Lys Ala Ser Lys Ser Leu
306 275 280 285

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/831,180

DATE: 05/23/2001

TIME: 16:02:15

Input Set : A:\50026.027001.SEQLIST.TXT

Output Set: C:\CRF3\05232001\I831180.raw

L:11 M:270 C: Current Application Number differs, Replaced Current Application No

L:11 M:271 C: Current Filing Date differs, Replaced Current Filing Date

L:179 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3